

# README

2026-05-13

## Overview

Readme for replicating “Beyond the Urban Sweet Spot: Firm-Level Evidence of Over-Agglomeration in Mongolia’s Capital” Working Paper Draft

This paper provides the first firm-level evidence on agglomeration economies in Mongolia, with a focus on Ulaanbaatar, the country’s dominant urban center. Using data from the 2021 Enterprise Census covering more than 17,000 firms, we estimate total factor productivity (TFP) and assess how it relates to localization, urban diversity, and peer productivity. The results reveal an inverted U-shaped relationship between industry clustering and productivity, indicating that agglomeration benefits peak and then decline as congestion sets in. These effects are especially pronounced in Ulaanbaatar and within manufacturing. While the cross-sectional nature of the data limits causal inference, the analysis provides novel insights into the spatial distribution of productivity and the efficiency costs of over-agglomeration. The findings underscore the importance of spatial planning, infrastructure investment, and regional diversification to sustain productivity growth in Mongolia’s increasingly concentrated economy.

For questions on the package please contact: Jose Luis Diaz Sanchez ([jdiazsanchez@worldbank.org](mailto:jdiazsanchez@worldbank.org)), Mijid-dorj Lkhagvasuren ([mlkhagvasuren1@worldbank.org](mailto:mlkhagvasuren1@worldbank.org))

## Data availability

Most data are publicly available, except the enterprise census Ulaanbaatar city’s district-level identifier, which was requested directly from the National Statistics Office (NSO) of Mongolia.

### 1. Enterprise Census of Mongolia

#### Enterprise census data (2021)

- File name: `data/Data_AANT2021.sav`
- Source: National Statistics Office (NSO) of Mongolia
- URL: <https://nada.nso.mn/mn/index.php/catalog/AANBT/> can be accessed upon registration (Mongolian language only)
- Access year: February 2025 (publicly available)

**Enterprise census Ulaanbaatar city’s district level identifier** The data were obtained directly from the National Statistical Office (NSO) to include Ulaanbaatar district-level identifiers. Ulaanbaatar is disaggregated into 9 districts (`Soum_code`), which are not available in the publicly released version.

- File name: `data/AANT1_UB_Dist_Soumcode.sav`
- Source: National Statistics Office (NSO) of Mongolia
- Access year: March 2025 (not public)

## 2. Administrative Boundaries (Shapefiles)

**Admin Boundaries level 0-2** Contains Mongolia’s administrative boundary shapefiles at three levels: Level 0 (country), Level 1 (aimag), and Level 2 (soum).

- File name: `data/mng-administrative-divisions-shapefiles/`
- Source: The Humanitarian Data Exchange
- URL: <https://data.humdata.org/dataset/cod-ab-mng>
- Access year: March 2025 (publicly available)

**Admin Boundaries level 3** Contains Mongolia’s administrative boundary shapefiles at level 3 (bag, khoroo)

- File name: `data/ulaanbaatar_mongolian.kmz`
- Source: Personal site Christopher M. Free, associate Researcher (research professor) at the Marine Science Institute and the Bren School of Environmental Science and Management at the University of California—Santa Barbara
- URL: <https://chrismfree.com/mongolia-gis-data/>
- Access year: March 2025 (publicly available)

**3. High Resolution Population Data** Provides population estimates at 30-meter spatial resolution.

- File name: `population_mng_2018-10-01`
- Source: AI for Good at Meta
- URL: <https://ai.meta.com/ai-for-good/datasets/high-resolution-population-density-maps/>
- Access year: March 2025 (publicly available)

**4. Time Use Survey** These datasets provide information on time allocation across economic and household activities of Mongolia.

### Mongolia Time Use Survey, 2011

- File name: `data/TUS11`
- Source: National Statistics Office (NSO) of Mongolia
- URL: <http://web.nso.mn/nada/index.php/catalog/90>
- Access year: March 2025 (publicly available)

### Mongolia Time Use Survey, 2023

- File name: `data/TUS23`
- Source: National Statistics Office (NSO) of Mongolia
- URL: <http://web.nso.mn/nada/index.php/catalog/192>
- Access year: March 2025 (publicly available)

**5. Ulaanbaatar’s Point of Interest data (firm’s location)** Points of interest (POI) is a geospatial dataset containing millions of real-world locations — including schools, businesses, hospitals, landmarks, and mountain peaks. Data downloaded covers Ulaanbaatar only

- File name: `data/ulaanbaatar.geojson`
- Source: Overture Maps
- URL: <https://docs.overturemaps.org/getting-data/>
- Access year: March 2025 (publicly available)

### Statement about Rights:

I certify that the author(s) of the manuscript have legitimate access to and permission to use the data used in this manuscript.

### Instructions for Replicators

In order to replicate the tables and figures in the paper:

- Run `replication_agglomeration.R`

Results (tables and figures) will be saved in the `output/` folder.

### List of exhibits

The provided code produces all tables and figures in the paper.

| Exhibit   | Output Name         | Script                                                     | Note                                                                                                                        |
|-----------|---------------------|------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------|
| Figure 1  | output/Figure1.xlsx | NA                                                         | The data, produced by the National Statistical Office (NSO), are provided in Excel format. The process is explained below.* |
| Figure 2A | output/Figure2A.png | Replication_agglomeration.R, scripts/Figure2_replication.R | Left panel of Figure 2: Illustrative purposes only; does not affect the main results of the paper.                          |
| Figure 2B | output/Figure2B.png | Replication_agglomeration.R, scripts/Figure2_replication.R | Right panel of Figure 2: Illustrative purposes only; does not affect the main results of the paper.                         |
| Figure 3  | output/Figure3.png  | Replication_agglomeration.R                                | This is for illustrative purposes only and does not affect the main results of the paper                                    |
| Figure 4  | output/Figure4.png  | Replication_agglomeration.R                                |                                                                                                                             |
| Table 1   | output/table1.pdf   | Replication_agglomeration.R, scripts/Table1_replication.R  |                                                                                                                             |
| Table 2   | output/table2.pdf   | Replication_agglomeration.R                                |                                                                                                                             |
| Table 3   | output/table3.txt   | Replication_agglomeration.R                                |                                                                                                                             |
| Table 4   | output/table4.txt   | Replication_agglomeration.R                                |                                                                                                                             |
| Table A1  | output/tableA1.txt  | Replication_agglomeration.R                                |                                                                                                                             |
| Table A2  | output/tableA2.txt  | Replication_agglomeration.R                                |                                                                                                                             |
| Table A3  | output/tableA3.txt  | Replication_agglomeration.R                                |                                                                                                                             |
| Table A4  | output/tableA4.txt  | Replication_agglomeration.R                                |                                                                                                                             |

**\*Figure 1 replication note:** Figure 1 is based on five series obtained from the National Statistics Office (NSO) of Mongolia (accessed via <https://1212.mn> in April 2025) and four intermediate calculations documented in the “Data” sheet of Figure1.xlsx. The final figure is located in the “Figure1” sheet.

Since official statistics on labor productivity for Ulaanbaatar are not available, we construct a proxy using publicly available data. The required inputs are:

1. GDP, current prices (national total)
2. GDP, current prices (Ulaanbaatar)
3. GDP, constant prices (base year 2010, national)
4. GDP, constant prices (base year 2015, national)
5. Employment (Ulaanbaatar)

Each step corresponds to a row in the “Data” sheet of Figure1.xlsx:

1. Since the constant price GDP series with base year 2015 starts only from 2015, it is linked to the 2010 base year series using growth rates. This produces a continuous series labeled “GDP, constant price 2015 (linked)”.
2. The GDP deflator is calculated by dividing GDP at current prices by GDP at constant prices (2015 linked).
3. Ulaanbaatar nominal GDP is deflated using the national GDP deflator, under the assumption that price dynamics in Ulaanbaatar are broadly similar to the national average, given that the capital accounts for approximately two-thirds of total GDP.
4. Real labor productivity in Ulaanbaatar is obtained by dividing real GDP by employment in Ulaanbaatar. This series is used as the final measure in Figure 1.

## Requirements

The replication requires Excel and R. Running under: macOS Ventura 13.3. The code was run on R 4.5.1 on a MacBook Air (M1, 8GB/256GB). It takes about 80 seconds to complete. The computational environment is fully reproducible using `renv`

## Code Descriptions

- `replication_agglomeration.R` This script loads the datasets, performs data cleaning, constructs the main variables, conducts the analysis, and generates tables and figures

This section includes auxiliary scripts for data processing and the production of illustrative figures and tables. These outputs are for illustrative purposes only and do not affect the main results of the paper.

- `scripts/Figure2_replication.R` Script that generates illustrative figures (Figure 2A and Figure 2B). It uses Meta population density data and Overture Maps place data to map the spatial distribution of residential areas and urban activity in Ulaanbaatar’s core districts.
- `scripts/Table1_replication.R` Script that produces Table 3. It uses the Time Use Survey to illustrate changes in commuting time between 2011 and 2023.

## Folder structure

```
library(fs)
```

```
## Warning: package 'fs' was built under R version 4.5.2
```

```
dir_tree(".",)
```

```
## .
## +-- Readme.Rmd
## +-- Readme.pdf
## +-- data
## |   +-- AANT1_UB_Dist_Soumcode.csv
## |   +-- Data_AANT2021.sav
## |   +-- TUS11
## |       +-- HD.sav
## |       +-- HH.sav
## |       +-- TUS 2011_DAIRY [mn].pdf
## |       +-- TUS 2011_HH [mn].pdf
## |       +-- TUS HL.sav
## |       \-- TUS HQ.sav
## |   +-- TUS23
## |       +-- dairy_main_1.dta
## |       +-- dairy_note_1.dta
## |       +-- hhdata_1.dta
## |       \-- ppdata_1.dta
## |   +-- mng-administrative-divisions-shapefiles
## |       +-- mng_admbnda_adm0_nso_20201019.CPG
## |       +-- mng_admbnda_adm0_nso_20201019.dbf
## |       +-- mng_admbnda_adm0_nso_20201019.prj
## |       +-- mng_admbnda_adm0_nso_20201019.sbn
## |       +-- mng_admbnda_adm0_nso_20201019.sbx
## |       +-- mng_admbnda_adm0_nso_20201019.shp
## |       +-- mng_admbnda_adm0_nso_20201019.shp.xml
## |       +-- mng_admbnda_adm0_nso_20201019.shx
## |       +-- mng_admbnda_adm1_nso_20201019.CPG
## |       +-- mng_admbnda_adm1_nso_20201019.dbf
## |       +-- mng_admbnda_adm1_nso_20201019.prj
## |       +-- mng_admbnda_adm1_nso_20201019.sbn
## |       +-- mng_admbnda_adm1_nso_20201019.sbx
## |       +-- mng_admbnda_adm1_nso_20201019.shp
## |       +-- mng_admbnda_adm1_nso_20201019.shp.xml
## |       +-- mng_admbnda_adm1_nso_20201019.shx
## |       +-- mng_admbnda_adm2_nso_20201019.CPG
## |       +-- mng_admbnda_adm2_nso_20201019.dbf
## |       +-- mng_admbnda_adm2_nso_20201019.prj
## |       +-- mng_admbnda_adm2_nso_20201019.sbn
## |       +-- mng_admbnda_adm2_nso_20201019.sbx
## |       +-- mng_admbnda_adm2_nso_20201019.shp
## |       +-- mng_admbnda_adm2_nso_20201019.shp.xml
## |       +-- mng_admbnda_adm2_nso_20201019.shx
## |       +-- mng_admbnda_region_nso_20201019.CPG
```

```

## | | +-- mng_admbnda_region_nso_20201019.dbf
## | | +-- mng_admbnda_region_nso_20201019.prj
## | | +-- mng_admbnda_region_nso_20201019.sbn
## | | +-- mng_admbnda_region_nso_20201019.sbx
## | | +-- mng_admbnda_region_nso_20201019.shp
## | | +-- mng_admbnda_region_nso_20201019.shp.xml
## | | +-- mng_admbnda_region_nso_20201019.shx
## | | +-- mng_admbndl_admALL_nso_itos_20201019.CPG
## | | +-- mng_admbndl_admALL_nso_itos_20201019.dbf
## | | +-- mng_admbndl_admALL_nso_itos_20201019.prj
## | | +-- mng_admbndl_admALL_nso_itos_20201019.sbn
## | | +-- mng_admbndl_admALL_nso_itos_20201019.sbx
## | | +-- mng_admbndl_admALL_nso_itos_20201019.shp
## | | +-- mng_admbndl_admALL_nso_itos_20201019.shp.xml
## | | +-- mng_admbndl_admALL_nso_itos_20201019.shx
## | | +-- mng_admbndl_region_nso_itos_20201019.CPG
## | | +-- mng_admbndl_region_nso_itos_20201019.dbf
## | | +-- mng_admbndl_region_nso_itos_20201019.prj
## | | +-- mng_admbndl_region_nso_itos_20201019.sbn
## | | +-- mng_admbndl_region_nso_itos_20201019.sbx
## | | +-- mng_admbndl_region_nso_itos_20201019.shp
## | | +-- mng_admbndl_region_nso_itos_20201019.shp.xml
## | | +-- mng_admbndl_region_nso_itos_20201019.shx
## | | +-- mng_admbndp_admALL_nso_itos_20201019.CPG
## | | +-- mng_admbndp_admALL_nso_itos_20201019.dbf
## | | +-- mng_admbndp_admALL_nso_itos_20201019.prj
## | | +-- mng_admbndp_admALL_nso_itos_20201019.sbn
## | | +-- mng_admbndp_admALL_nso_itos_20201019.sbx
## | | +-- mng_admbndp_admALL_nso_itos_20201019.shp
## | | +-- mng_admbndp_admALL_nso_itos_20201019.shp.xml
## | | \-- mng_admbndp_admALL_nso_itos_20201019.shx
## | +-- population_mng_2018-10-01.csv
## | +-- ulaanbaatar.geojson
## | \-- ulaanbaatar_mongolian.kmz
## +-- output
## | +-- Figure1.xlsx
## | +-- Figure2A.png
## | +-- Figure2B.png
## | +-- Figure3.png
## | +-- Figure4.png
## | +-- table1.pdf
## | +-- table2.pdf
## | +-- table3.txt
## | +-- table4.txt
## | +-- tableA1.txt
## | +-- tableA2.txt
## | +-- tableA3.txt
## | \-- tableA4.txt
## +-- renv.lock
## +-- replication_agglomeration.R
## \-- scripts
##   +-- Figure2_replication.R
##   \-- Table1_replication.R

```